# PD8 Exh A

# Data Appendix

**March 25, 2019**

# Table of Contents

# CDC Mortality Data

**Mortality Data Sources**

The National Center for Health Statistics (NCHS), part of the Centers for Disease Control, constructs Multiple Cause of Death (MCOD) data.[1] The data are individual-specific, providing information on the individual's cause of death, and are based on death certificates.

Public MCOD data are available by year for download in machine-readable format from 1968-2017; however, these data contain some key limitations with respect to the geographic detail provided:

- From 1989-2004, the data mask county identifiers for counties with population under 100,000.
- From 2005-2017, the data omit all geographic details including state.[2]

To supplement the publicly available data, we received restricted-use MCOD data for the years 2005-2016 which include all county- and state-level identifiers.[3] These data omit the day of the week and month of death.

Because of the limitations in geographic detail in the public 2017 data file, most of our analysis will concentrate on data through 2016, the last year for which we have complete county and state identifiers.

**Cause of Death and Identifying Opioid-Related Overdose Mortality**

The NCHS data include codes based on the World Health Organization's International Classification of Diseases (ICD) that can be used to identify whether a death was due to opioid overdoses.[4] This is a system of codes frequently used in health care and public health fields to standardize diagnosis and treatment codes across data sources and counties.[5]

This system of codes is periodically revised. For the years 1999 through 2017, the NCHS mortality data incorporate the codes as defined by the 10th revision of ICD (known as ICD-10). From 1979 through 1998, the NCHS mortality data incorporate the 9th revision of ICD (known as ICD-9). As discussed in more detail below, we account for the transition between these code systems as necessary.

Each record in the NCHS mortality data contains two forms of ICD codes:

- A single code for the "Underlying Cause of Death".

---

[1] https://www.cdc.gov/nchs/nvss/deaths.htm

[2] We accessed the data through the National Bureau of Economic Research's SAS-formatted datasets, available here: https://www.nber.org/data/vital-statistics-mortality-data-multiple-cause-of-death.html

[3] These data have been made available based on completion of a data use agreement with the NCHS and are subject to confidentiality restrictions which forbid (among others) the reporting of statistics for a county or state based on fewer than 10 deaths.

[4] http://www.who.int/health-topics/international-classification-of-diseases

[5] In addition to use in classification of causes of death in mortality data, the ICD system is also used in coding and classifying diagnosis and morbidity data at inpatient and outpatient medical facilities through the ICD-10 CM (clinical modification) system. https://www.cdc.gov/nchs/icd/index.htm. For background on the development and accuracy of the ICD code system, see: O'Malley, Kimberly J et al. "Measuring diagnoses: ICD code accuracy" Health services research vol. 40,5 Pt 2 (2005): 1620-39.

- Up to 20 additional codes for the "Multiple Causes of Death"

Based on these two sets of codes, we identified all deaths in a given year with codes that indicate that opioids are identified as a cause of death. In doing so, we follow the procedures developed by the CDC to identify and classify drug overdoses as opioid-related:

- The CDC's "Annual Surveillance Report of Drug-Related Risks and Outcomes" for the year 2018outlines the codes used by the CDC under ICD-10.[6]
- "The Numbers Behind the Opioid Crisis", a 2017 report prepared by the Social Capital Project for the Joint Economic Committee, outlines the codes used by the CDC under ICD-9.[7]

For ICD-10, we further categorize opioid-related drug overdoses into one of three categories:

- Prescription Opioid overdoses only
- Any Heroin overdoses, excluding those with Fentanyl
- Any Fentanyl overdoses, with any additional opioids

The codes used in our analysis are identified in Table 1 below.

[6] https://www.cdc.gov/drugoverdose/pdf/pubs/2018-cdc-drug-surveillance-report.pdf
[7] https://www.lee.senate.gov/public/index.cfm/2017/10/the-numbers-behind-the-opioid-crisis

**Table 1**
**ICD-9 and ICD-10 Codes Used to Identify Opioid-Related Overdose Mortality**

| | ICD-10 (1999-2016) | | ICD-9 (1983-1998) | |
|---|---|---|---|---|
| **Category** | **Underlying Cause** | **Contributing Cause** | **External Cause of Injury** | **Diagnosis** |
| Any Opioid Mortality | X40 X41 X42 X43 X44 X60 X61 X62 X63 X64 X85 Y10 Y11 Y12 Y13 Y14 | T40.0 T40.1 T40.2 T40.3 T40.4 T40.6 | E850 E851 E852 E853 E854 E855 E856 E857 E858 E950.0 E950.1 E950.2 E950.3 E950.4 E950.5 E962.0 E980.0 E980.1 E980.2 E980.3 E980.4 E980.5 | 965.00 965.01 965.02 965.09 |
| Detailed Opioid Type: | | | | |
| Heroin | X40 X41 X42 X43 X44 X60 X61 X62 X63 X64 X85 Y10 Y11 Y12 Y13 Y14 | T40.1 | N/A | |
| Prescription Opioids | | T40.2 | | |
| Methadone | | T40.3 | | |
| Fentanyl | | T40.4 | | |
| Other and Unidentified Opioid | | T40.6 | | |
| | | | | |
| Unidentified Drugs | X40 X41 X42 X43 X44 X60 X61 X62 X63 X64 X85 Y10 Y11 Y12 Y13 Y14 | T50.9 | E850 E851 E852 E853 E854 E855 E856 E857 E858 E950.0 E950.1 E950.2 E950.3 E950.4 E950.5 E962.0 E980.0 E980.1 E980.2 E980.3 E980.4 E980.5 | 965.9 977.9 |

**Allocating Drug Overdoses with Unspecified Drug Types**

There is a sizeable share of deaths that are due to drug overdoses where the identification of the drug(s) contributing to the death are not identified in the MCOD data. For these deaths, we apply a methodology developed by Christopher Ruhm in a series of papers in 2017 and 2018 to probabilistically allocate these unspecified drug overdoses to opioids and other causes based on the additional characteristics included in the underlying data.[8] We have implemented the methodology outlined by Ruhm in order to allocate these overdoses to either opioids or non-opioids.

This approach uses a logistic regression to evaluate the relationship between demographic factors and whether a death was opioid-related based on individuals for whom the source of an overdose death was

[8] Ruhm, Christopher J., "Geographic Variation in Opioid and Heroin Involved Drug Poisoning Mortality Rates." American Journal of Preventive Medicine 53, No. 6 (2017): 745-753; Ruhm, Christopher J., "Corrected US opioid-involved drug poisoning deaths and mortality rates, 1999–2015," Addiction 113 (2018): 1339-1344.; Ruhm, Christopher J., "Deaths of Despair or Drug Problems?" NBER Working Paper 24188, January 2018.

reported. Logistic regressions were run on a year-specific basis for the years 1983-2016. After running these regressions, national and county-level time series of opioid overdose deaths were constructed.

Figure 1 below plots overdose deaths where the cause was identified as opioid-related as well as overdose deaths where the drug involved was not specified. Figure 2 plots the results of the Ruhm Adjustment to allocate the unspecified drug deaths.

**Figure 1**
**Opioid vs. Unspecified Overdose Deaths**
**National 1983-2016**

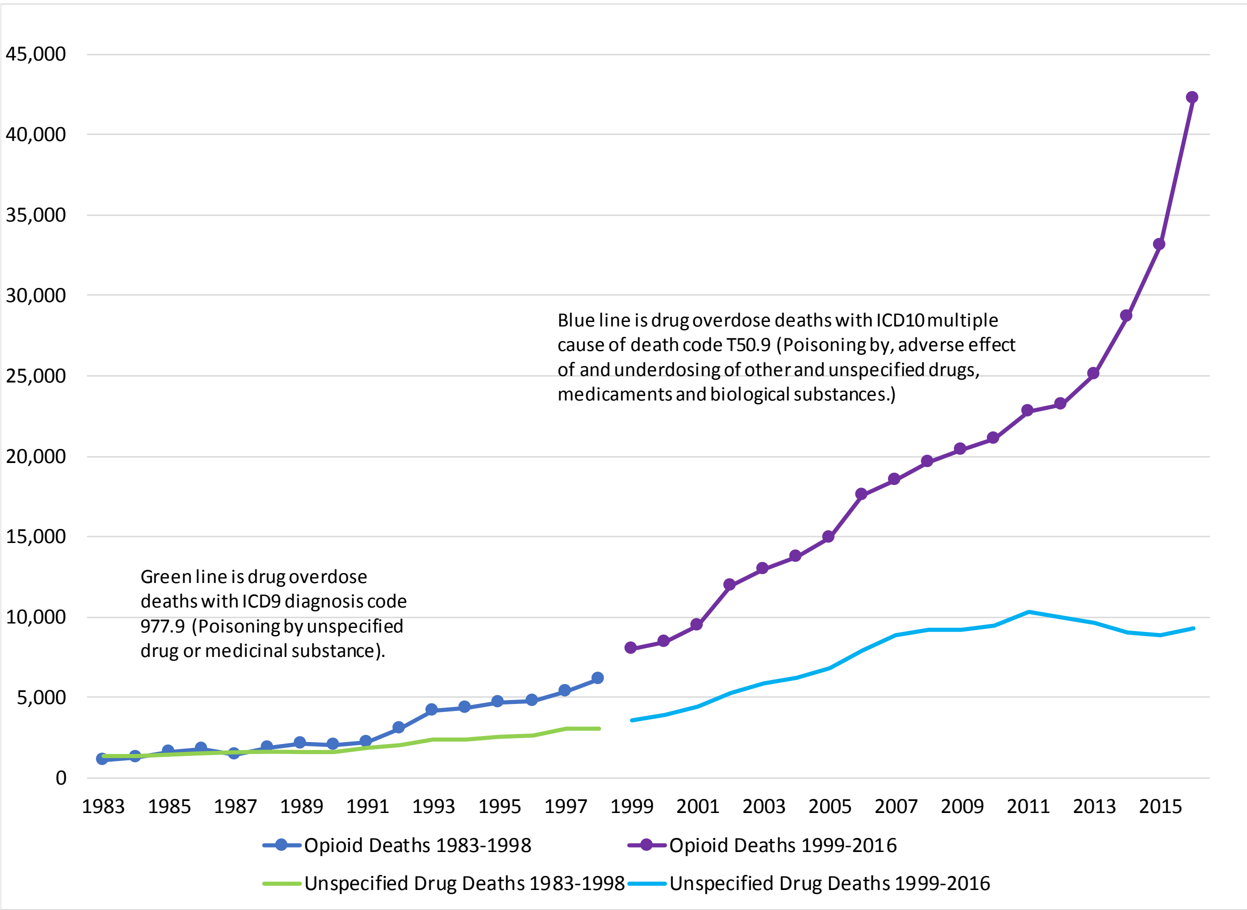
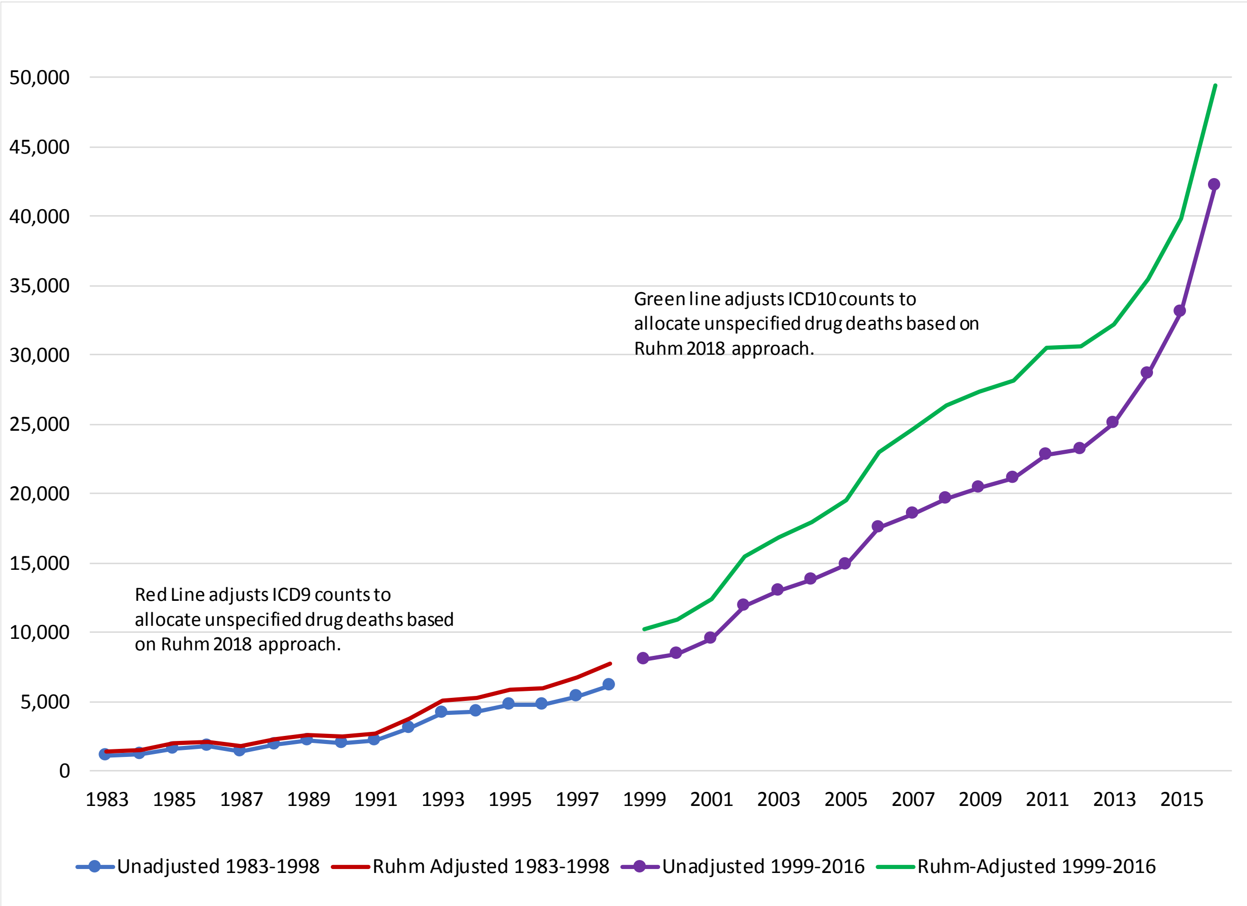


**Figure 2**
**Impact of Ruhm Adjustment on Opioid Overdose Deaths**
**National 1983-2016**




In addition, there are a sizeable share of drug overdoses which are identified as opioid-related but for which the data do not specify whether they are due to prescription opioids, heroin or other substances. We follow the procedures outlined in Ruhm 2018 to further allocate opioid overdoses where the type of opioid involved is not reported in the MCOD data into the following categories:

- Prescription Opioid overdoses only
- Any Heroin overdoses, excluding those with Fentanyl
- Any Fentanyl overdoses, with any additional opioids

This approach also uses a logistic regression to evaluate the relationship between demographic factors and the type of opioid, limited to mortalities for which the detailed opioid type involved was reported. Regressions were run on a year-specific basis for the years 1999-2016. After performing these regressions, national and county-level time series for mortality rates for each of the three detailed opioid categories were constructed.

**Adjusting for the ICD9 to ICD10 transition**

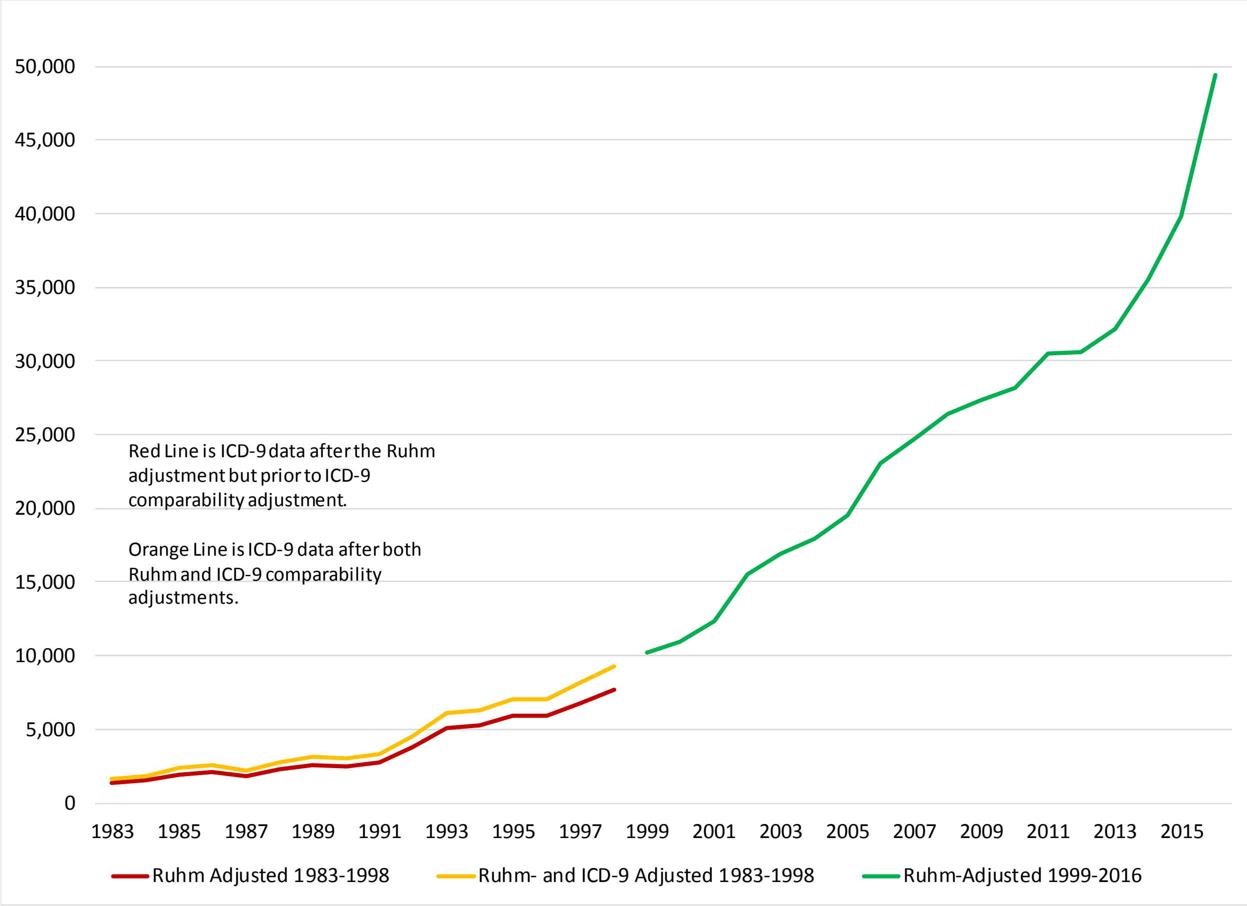
The ICD code system is periodically updated. The CDC warns that when comparing mortality rates across regimes, differences in the coding regimes can bias the analysis of time series trends.[9]

In order to account for the transition from ICD-9 to ICD-10, we adjust the ICD-9 mortality rates by the Comparability Ratio calculated in Hoyert et al. 2001.[10] This ratio was based on the review of a subset of approximately 80% of death records which were categorized separately under both ICD9 and ICD 10. Based on this sample, the authors developed a comparability ratio for "Drug-Induced Deaths" of 1.195, indicating that mortality rates under ICD-9 should be adjusted upward by about 20% in order to be comparable to rates under ICD-10.[11]

The impact of the adjustment of ICD-9 data is shown in Figure 3, which demonstrates continuity between the adjusted data from the ICD-9 regime and the ICD-10 regime.

[9] Anderson, Robert N., Arialdi M. Minino, Donna L. Hoyert, and Harry M. Rosenberg, "Comparability of Cause of Death Between ICD-9 and ICD-10: Preliminary Estimates," National Vital Statistics Report 49, No. 2 (2001): 1-32.
[10] Donna L. Hoyert, Elizabeth Arias, Betty L. Smith, Sherry L. Murphy, Kenneth D. Kochanek, "Deaths: Final Data for 1999," National Vital Statistics Reports 49 No. 8 (2001): 1-114.
[11] Note that the category "Drug-Induced Deaths" is broader than deaths due to drug overdoses.

**Figure 3**
**Opioid Overdose Deaths with and without ICD-9 Adjustment**
**National, 1983-2016**




**Constructing age-adjusted mortality rates**

In analyzing mortality data, researchers frequently focus on age-adjusted mortality rates. This allows for comparison across geographies and across years which control for changes over time in the age distribution of the population as well as differences in the age distribution of the local population for mortality rates constructed for different geographic areas.

We have followed the practice outlined in Murphy et al. (2017) to construct age adjusted opioid mortality rates for each county and year analyzed.[12] This approach relies on reweighting the opioid mortality rates for each county and year by applying weights by age category from the standardized national US population for the year 2000. The impact of this age-adjustment on mortality rates are minor.

[12] Sherry L. Murphy, Jiaquan Xu, Kenneth D. Kochanek, Sally C. Curtin, and Elizabeth Arias, "Deaths: Final Data for 2015," National Vital Statistics Reports 66 No. 6 (November 27, 2017): 1-75.

**Sub-setting the data for analysis**

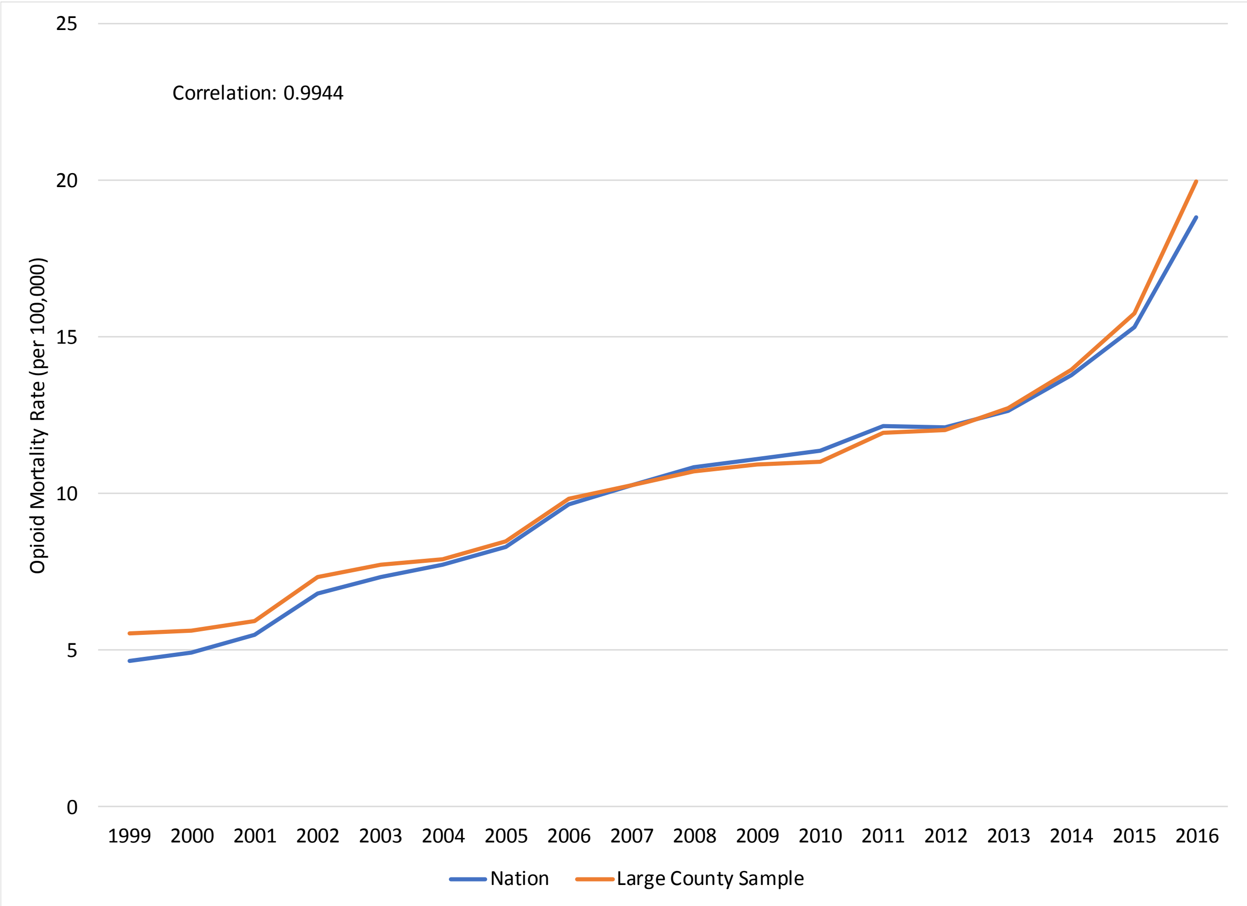
Some of the analyses of MCOD data in the expert reports are based on a subset of the data which includes "large counties". These are the unmasked counties identified explicitly in the MCOD data over the period of 1993-2004, and which are thus defined as "large counties" by the NCHS. While the restricted use data are available for all counties from 2005-2016, the opioid crisis was well underway by 2005 and in order to estimate the relationship between the rise of shipments and mortality, a longer time series is necessary.

As noted above, county identifiers are masked in the MCOD data for small counties (defined as population of less than approximately 100,000) and thus are not reported by NCHS for the years 1989-2004.[13] The analysis sample includes 404 counties in total, which together account for 69.1% of US adult (15+) population in 2016. We rely on the publicly available data from 1993-2004 and the restricted use data from 2005-2016.

Figure 4 below compares the opioid mortality rate for the large county sample to the nation as a whole for 1999 to 2016. The correlation between these series is .9944.

[13] The CDC used Census population to determine whether a county meets the 100,000 threshold. From 1989-1993, the CDC applied the 1980 Census population; from 1994-2002, the CDC applied the 1990 Census population; from 2003-2004, the CDC applied the 2000 Census. See: https://www.cdc.gov/nchs/nvss/mortality_public_use_data.htm.

**Figure 4**
**National and Large County Sample Opioid Mortality Rates**
**1999-2016**



## ARCOS Prescription Shipment Data

The Drug Enforcement Agency (DEA) provides data on shipments of prescription opioids over time and across geographies. This appendix describes the source of these data and the steps taken to process and set up the data for analysis.

**DEA ARCOS Data**

The Controlled Substances Act of 1970 requires all manufacturers and distributors of controlled substances to regularly report the details of their transactions of controlled substances to the federal government. The DEA is responsible for managing and organizing this reporting process and sharing the data with investigative and regulatory agencies to monitor diversion of controlled substances to illicit markets.[14]

The DEA has developed a data collection system to automate this reporting process known as the Automation of Reports and Consolidated Orders System ("ARCOS"). The DEA releases publicly available annual reports summarizing the ARCOS data.[15] These public reports, known as Retail Drug Summary Reports, identify the total grams of each controlled substance shipped to retail registrants, excluding intermediary shipments from manufacturer to distributor or between distribution centers. Table 1 in these public reports summarizes shipments at the most detailed level available: by 3-digit zip code, by DEA drug code, and by quarter-year.

The DEA includes the following opioid drug codes in the ARCOS data[16]:

| DEA Drug Code | Drug Name |
|---|---|
| 9050 | CODEINE |
| 9064 | BUPRENORPHINE |
| 9120 | DIHYDROCODEINE |
| 9143 | OXYCODONE |
| 9150 | HYDROMORPHONE |
| 9193 | HYDROCODONE |
| 9230 | MEPERIDINE (PETHIDINE) |
| 9250B | METHADONE |
| 9300 | MORPHINE |
| 9639 | OPIUM POWDERED |
| 9652 | OXYMORPHONE |
| 9780 | TAPENTADOL |
| 9801 | FENTANYL BASE |

[14] https://www.deadiversion.usdoj.gov/arcos/index.html
[15] https://www.deadiversion.usdoj.gov/arcos/retail_drug_summary/index.html
[16] The data also include non-opioid controlled substances such as Amphetamine.

DEA drug codes provide the molecule of the controlled substance, but not the dosage size, brand name, or format of the substance (i.e., tablet, transdermal patch, or solution). For example, shipments of OxyContin would be classified under the drug code 9143 Oxycodone along with all other branded and generic opioids containing Oxycodone. In analyzing the ARCOS data, we focus on opioids typically used for treatment of pain and not for medially-assisted treatment of use disorder. We thus exclude shipments of Buprenorphine (Drug Code 9064) and Methadone (Drug Code 9250B).[17]

The DEA website includes public reports in PDF format covering the years 2000-2017. We processed these reports, extracted the data from Table 1, and converted the underlying data into machine readable format. Additional reports covering the years 1997, 1998, and 1999 were identified based on an archived version of the DEA website available through the Internet Archive.[18] We processed these files in the same way.

**DEA Schedule**

Under the Controlled Substances Act, the DEA classifies drugs into five categories based on their medical purpose and the potential for abuse or dependency. Schedule I drugs, for example, include heroin and other illegal narcotics that are not accepted for medical use. Schedule II drugs are those with "high potential for abuse, with use leading to severe psychological or physical dependence."[19] Schedule III drugs have moderate to low potential for abuse or dependence. Most prescription opioids are classified by the DEA as Schedule II, including OxyContin. However, products containing fewer than 90 milligrams of codeine (typically in combination with Tylenol) are classified as Schedule III.

The publicly available ARCOS data do not break out shipments by DEA Schedule. It is thus not possible to limit our analysis of shipments to Schedule II opioids. The two drug codes which are likely to contain meaningful shipments of Schedule II drugs (9050 – Codeine, 9120 – Dihydrocodeine) are a relatively small share of total shipments (fewer than 2.5% of total shipments from 1997-2010).

**Morphine Milligram Equivalents (MMEs)**

In order to compare shipments of different opioid types, it is necessary to standardize the volumes to account for differences in the strength of the underlying molecule. The typical way to standardize opioids is to convert grams of the underlying substance into Morphine Milligram Equivalents (MMEs). Conversion to MME is common practice in the literature analyzing prescription opioids. For example,

---

[17] While there are forms of Buprenorphine and Methadone that are prescribed for the treatment of pain, most shipments of these drugs are for use in treatment of use disorder. The ARCOS public reports do not permit disaggregation by use case; however, the data do report total national shipments in grams by "business activity" including pharmacies and narcotic treatment programs which dispense Methadone for treatment of use disorder. In 2010, there were a total of 8.7 million grams of Methadone shipped to such narcotic treatment programs, 44% more than total grams of Methadone shipped to pharmacies. https://www.deadiversion.usdoj.gov/arcos/retail_drug_summary/2010/2010_rpt7.pdf.

[18] https://web.archive.org/web/20090320043126/https://www.deadiversion.usdoj.gov/arcos/retail_drug_summary/index.html

[19] https://www.dea.gov/drug-scheduling.

the CDC's prescribing guidelines for prescription opioids are expressed in MMEs.[20] Economic literature analyzing the links between prescription shipments and harms typically convert ARCOS shipments to MMEs.[21]

In order to convert the ARCOS shipments data to MMEs, we rely on conversion factors commonly used in the literature.[22] By multiplying shipments in grams by the conversion factor and dividing by 1000, we express the shipments in MMEs. The conversion factors we use are:

| DEA Drug Code | Drug Name | MME Conversion Factor |
|---|---|---|
| 9050 | CODEINE | 0.15 |
| 9120 | DIHYDROCODEINE | 0.25 |
| 9143 | OXYCODONE | 1.50 |
| 9150 | HYDROMORPHONE | 4.00 |
| 9193 | HYDROCODONE | 1.00 |
| 9230 | MEPERIDINE (PETHIDINE) | 0.10 |
| 9300 | MORPHINE | 1.00 |
| 9639 | OPIUM POWDERED | 1.00 |
| 9652 | OXYMORPHONE | 3.00 |
| 9780 | TAPENTADOL | 0.40 |
| 9801 | FENTANYL BASE[23] | 100.00 |

**Mapping Shipments from 3-Digit Zip Codes to Counties**

As noted above, the most detailed geographic area reported in the public ARCOS reports is the 3-digit zip code. 3-digit zip codes are based on the first three digits of standard US postal zip codes. These areas typically (but not exclusively) span across more than one county and thus are not directly comparable to the county-level data available for mortality, crime, and demographic and economic statistics.

[20] https://www.cdc.gov/mmwr/volumes/65/rr/rr6501e1.htm

[21] *See*: Abby Alpert, David Powell, and Rosalie Liccardo Pacula, "Supply-Side Drug Policy In The Presence Of Substitutes: Evidence From The Introduction Of Abuse-Deterrent Opioids," American Economic Journal: Economic Policy, 2018, 10(4): 1-35; Christopher J. Ruhm, "Deaths of Despair or Drug Problems?" NBER Working Paper 24188, January 2018.

[22] https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/Opioid-Morphine-EQ-Conversion-Factors-Aug-2017.pdf

[23] MME factors for Fentanyl used by medical practitioners depend in part on the transmittal mechanism, as prescription Fentanyl is typically contained in transdermal patches or lozenges. The DEA ARCOS data aggregates shipments of Fentanyl across types into total grams of the base drug. In order to convert these total grams of Fentanyl to MMEs, we apply a factor of 100, consistent with the literature which applies the same factor to convert parenteral Fentanyl to morphine equivalents. See: footnote 8 from: https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/Opioid-Morphine-EQ-Conversion-Factors-Aug-2017.pdf

In order to link the ARCOS shipments data to the other county data, we have allocated shipments based on the weighted average population of census block centroids (center points) that fall within each county that a 3-digit zip code crosses. This means that when a 3-digit zip code crosses county boundaries, we use the population at the census block level to estimate the share of population across counties for the 3-digit zip. An underlying assumption to this approach is that the shipments per capita within a 3-digit zip code are the same across census blocks. This is a standard method used to perform this type of geographic allocation.[24]

Census block centroids are useful as a translation unit because both 3-digit zips and counties are large compared to census blocks and county borders perfectly match census geographies. Additionally, since drug shipments are sold to end-users, population weighting is preferable to area weighting.

To do this, we first determine which census blocks fall within each 3-digit zip. Then we determine which counties each of those blocks belong to and compute the share of the 3-digit zip's population that falls within that county, based on 2010 population from the US Census. Finally, we allocate the ARCOS drug shipment totals to those counties according to their population share.

$$Qc_i = \sum_{j=1}^{n} Qz_j \frac{Pc_{ij}}{Pz_j}$$

Given:

$$\begin{aligned} Qc_i &= Quantity\ in\ county\ i \\ Qz_i &= Quantity\ in\ zip\ j \\ Pc_{ij} &= Population\ in\ county\ i\ and\ zip\ j \\ Pz_j &= Population\ in\ zip\ j \end{aligned}$$

After this allocation, we have a dataset with county-level shipments by drug code and quarter-year that can be combined with other county-level information used in our analysis.

**Interpolation for Year 2000**

The DEA public report for calendar year 2000 contains shipments for only a subset of the drug codes used in the other years. While oxycodone and hydrocodone are reported, remaining opioids including morphine are not. In order to account for this change in reporting and to create a consistent time series, we have interpolated shipments for the omitted drug codes at the county level.

---

[24] For example, Rolheiser et al. (2018) perform a similar allocation to convert county-level prescribing rates to Congressional districts based on census block population weights. Lyndsey A. Rolheiser, Jack Cordes, BSPH, S.V. Subramanian, "Opioid Prescribing Rates by Congressional Districts, United States, 2016,". American Journal of Public Health, 108, no. 9 (September 1, 2018): pp. 1214-1219.

For each county and omitted drug code, we interpolate shipments in the year 2000 as the midpoint between shipments in the years 1999 and 2001.[25]

**Shipments per Capita per Day**

In order to account for differences in population across counties, we convert total MMEs into MMEs per adult by dividing by US Census data reporting adult population (aged 15 plus) for each county and year in the data. Finally, in order to express the units in easier-to-interpret scales, we divide the per adult shipments by 365 to generate MMEs per capita per day. This is the unit of shipments that we focus on in our analysis and can be interpreted as the amount of opioids shipped to a county in a given year per adult and per day. To match common nomenclature, we refer to this as shipments per capita per day.

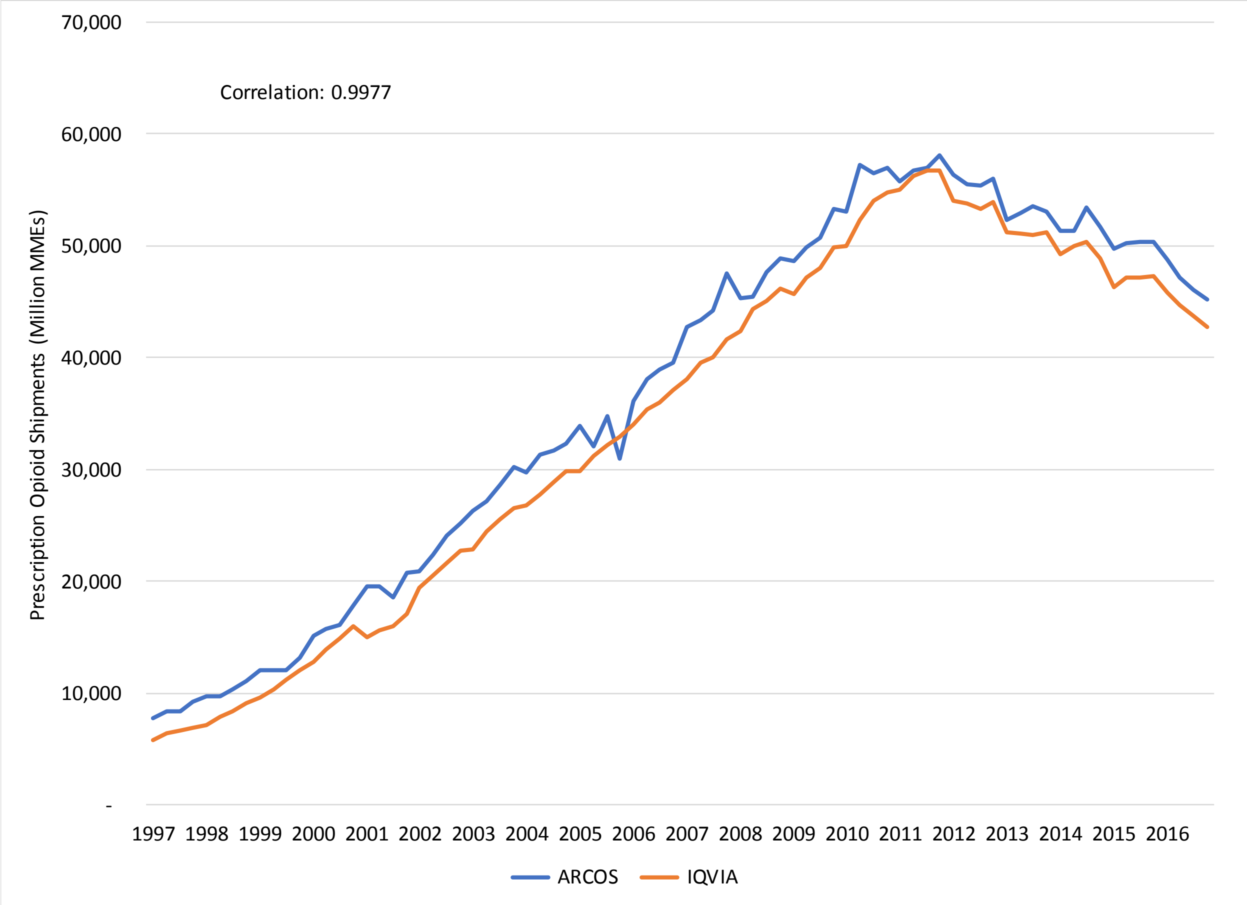
**Comparison between ARCOS and IQVIA Data**

Because the ARCOS data provides geographic variation in shipments across areas, many of the analyses rely on the ARCOS data. Some of the analysis in the Rosenthal Report, however, relies on an alternative data source known as IQVIA to estimate the relationship between marketing and sales of prescription opioids. As described in more detail in the Rosenthal Report, IQVIA data provide both marketing and sales information on prescription opioids. Unlike the ARCOS data, the IQVIA data provides sales of prescription opioids by detailed NDC code, which allows disaggregation by brand name, dosage size, transmittal mechanism, and DEA Schedule (among others). As a result, the analyses relying on IQVIA data are limited to shipments of Schedule II opioids along with shipments of Butrans (a transdermal form of Buprenorphine typically prescribed for treatment of pain).

Despite the differences in the how these two data sources are constructed, they are very comparable. Figure 5 below compares quarterly shipments of opioids in MMEs from the ARCOS and IQVIA data. The correlation between these two time series is 0.9977.

[25] According to national opioid shipment data provided by the International Narcotics Control Board, shipments in MMEs per capita in the US in the year 2000 were close to the mid-point between 1999 and 2001. The midpoint of 1999 and 2001 from these data (excluding Methadone) is 171.98, compared to actual value of 171.82. http://www.painpolicy.wisc.edu/sites/default/files/country_files/morphine_equivalence_wo_methadone/unitedstatesofamerica_me.pdf

**Figure 5**
**ARCOS vs. IQVIA Prescription Opioid Shipments**
**1997-2016, Morphine Milligram Equivalents per Quarter**



## UCR Crime Data

**UCR Data Construction**

The FBI Crime by County data compiles the number of Part I crimes[26] brought to the attention of Law Enforcement Agencies (LEAs).[27] These data have been collected and submitted by the Uniform Crime Reporting (UCR) Programs at reporting agencies[28] and are available by LEA by county by year from 1980 to 2016.[29] LEAs report on a voluntary basis,[30] and as such reporting by LEAs in a given county can vary over time.[31] The UCR data also reports the population of the jurisdiction that is covered by that LEA (referred to as "covered population").[32]

In order to merge these data with economic and demographic variables measured at the county level, Federal Information Processing Standard (FIPS) county codes are assigned to the data using a FIPS/LEA mapping file created by the National Archive of Criminal Justice (NACJD).[33] This crosswalk provides a unique FIPS code for each reporting LEA. An LEA can have jurisdiction across multiple counties and the UCR data reports crime counts for that LEA in each of the different counties.[34] However, the FIPS/LEA mapping file only includes the FIPS code for the primary county for that LEA. As a result, there are instances in which the incorrect "primary" FIPS code could be mapped to that LEA's other (secondary) UCR County, creating a mapping in which a single UCR county would be associated with multiple FIPS. To account for this issue, two rules were implemented to identify and remove the improperly matched FIPS code to ensure that each UCR county was assigned a unique FIPS code. First, any FIPS code assigned to a UCR county that was associated with an LEA with no reported population was dropped.

---

[26] Part I Crimes are broken into violent crimes and property crimes and include: criminal homicide, rape, robbery, aggravated assault, burglary, larceny-theft, motor vehicle theft, and arson. The data do not report incidents of crimes against society (e.g. drugs, prostitution).

[27] Use of the UCR data in constructing county crime rates have been discussed in prior economic studies including Dave, Dhaval, Monica Deza, and Brady P. Horn. Prescription Drug Monitoring Programs, Opioid Abuse, and Crime. No. w24975. National Bureau of Economic Research, 2018; Maltz, Michael D., and Joseph Targonski. "A note on the use of county-level UCR data." Journal of Quantitative Criminology 18, no. 3 (2002): 297-318; and Lott, John R., and John Whitley. "Measurement error in county-level UCR data." Journal of Quantitative Criminology 19, no. 2 (2003): 185-198.

[28] For details on the FBI's Uniform Crime Reporting Methodology, see https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/resource-pages/methodology. Also see https://ucr.fbi.gov/nibrs/2016/resource-pages/methodology-2016_final_.pdf.

[29] Note that there were two changes to the UCR data in 1994 including a change to the algorithm used to adjust for incomplete reporting and a change in the population coverage variable provided (see United States Department of Justice Federal Bureau of Investigation 'Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1996', p. 7). As a result, the data are restricted to 1995 and later.

[30] Note, however, that "if agencies with populations of 100,000 or more are missing reports, they are contacted by FBI personnel and urged to complete their reports." https://www.ncjrs.gov/pdffiles1/nij/grants/215343.pdf p. 11. See FBI CJIS Division, UCR Program, "Summary Reporting System (SRS) User Manual," June 20, 2013, available at https://ucr.fbi.gov/nibrs/summary-reporting-system-srs-user-manual, pp. 22-23.

[31] If an LEA has jurisdiction that covers more than one county, the FBI allocates the reporting data from one LEA across the relevant counties.

[32] Note that the reported population accounts for any overlapping jurisdictions in order to avoid double-counting of population. For example, some reporting LEAs are university campus police. For these LEAs, UCR reports a covered population of 0 since the jurisdiction's population would already be counted in that city/county's police force.

[33] Bureau of Justice Statistics. "Law enforcement agency identifiers crosswalk, 2012." (2013).

[34] For example, the Atlanta Police Department is active in two counties, and will appear twice in a given year in the UCR offense data – once for each county.

Second, FIPS codes that met all of the following criteria were flagged as likely improperly assigned and dropped: (1) not the most common FIPS code associated with a UCR county; (2) was associated with a covered population that accounted for less than 5% of the UCR county; and (3) was assigned to fewer than 20% of the reporting LEAs for that UCR county. Finally, any remaining UCR counties that were still associated with more than one FIPS code were manually reviewed to determine the appropriate one-to-one FIPS code to UCR county mapping. Finally, some additional manual adjustments were made to certain FIPS variables to account for cities with their own FIPS codes and FIPS code changes.[35]

Certain reporting LEA observations were then modified due to data inconsistencies or inconsistent geographic designation over time.

- Selected state police LEAs were removed as the population coverage suggested enforcement in multiple counties– these include: "CTCSP00" for Connecticut; "VTVSP00" for Vermont; "AKASP00" and "AKAST01" for Alaska.
- In 2001, Colorado created a new county, Broomfield County, out of portions of four neighboring counties (Weld County, Jefferson County, Boulder County, and Adams County). As a result of this shift in geographic boundaries over time, Broomfield County (after 2001) and the other four counties are dropped from the sample.
- Two counties in VA (Halifax County and Allegheny County) were also dropped due to boundary changes in these counties during the 1995 – 2016 period.

The UCR data does not distinguish between when crime reported is missing or equal to 0. Such missing information is interpreted as equal to 0 if that LEA reported a positive value for another type of crime. For example, if an LEA has a missing value for property crime but reported a positive value for violent crime in that year, the property crime missing is assumed to reflect a property crime count of 0. If both counts are missing for that LEA in that year, the LEA is assumed not to have reported crime counts for that year (assign the value as a true missing). Finally, the UCR data provides information on the number of months (in a reporting year) for which the crime data are reported. If an LEA reported crime for a period of less than 12 months, the annual count of crime offenses is imputed by scaling the counts by 12/months reported.[36] LEA-years with fewer than 2 months reported are dropped from the sample. Crime rates are then calculated as the sum of crime offenses (in that crime category) across the county divided by the total covered population across the reporting LEAs. Note that this population may be, and often is, less than the total county population as there are some counties in which not every LEA in that county has reported.

The analysis is based on counties with an average (1995/1996) census population of over 100,000 and to those counties in which the reporting LEA covered population accounts for at least 85% of the county population in at least one year of each period of interest (1995/1996 and 2015/2016). These data were

---

[35] Independent cities were coded to their FIPS (Baltimore, St. Louis and independent cities in Virginia). Washington DC and selected counties (Anchorage AK, Davidson County TN, and Arlington VA) without a FIPS code from the NACJD mapping process were adjusted manually. New York City was coded as being Manhattan because the FBI allocates all crime incidents occurring in any borough of New York City to Manhattan. Some FIPS codes were consolidated according to http://www.nber.org/asg/ASG_release/County_City/FIPS/FIPS_Changes.pdf.
[36] This is similar to the method used by the NACJD, see https://www.icpsr.umich.edu/icpsrweb/content/NACJD/guides/ucr.html

then manually reviewed to identify and drop outlier observations that appear to be reflective of data errors or systematic changes in reporting.[37]

Analysis is based on the average crime rate for each of the two periods analyzed: 1995/96 and 2015/2016.[38] A "pre-period" county crime rate is calculated for both property crime and violent crime as the average of that county's rate in 1995 and 1996.[39] A "post-period" crime rate is calculated as the average across 2015 and 2016 crime rates. The dependent variable in the crime regression is calculated as the difference between these two numbers.

The dataset used for figures presented in the Gruber Report and the crime regression model presented in the Cutler Report includes 417 counties. There are 487 counties in the UCR data with 1995-1996 census population of at least 100,000; of these 487 counties, 66 are dropped due to having no pre- or post-period crime rate due to unavailability of data or insufficient population coverage ; another 4 counties are dropped due to having high positive outliers for shipments per day, likely reflective of being transshipment counties (1997-2010 shipments per capita per day greater than 3.5).[40]

---

[37] Examples include (1) changes over time in an LEA's population that appear to reflect jurisdiction changes but for which the crime reporting does not appear to also adjust (thus yielding a significant change in crime rate in a particular year); or (2) significant one-time changes in crime counts that suggest reporting errors.

[38] Some observations include only one year of the pre-period or post-period due to the 85% population threshold not being met. In this case, only one year is used to calculate the pre- or post-period crime rate.

[39] Note that because there was a change in the method of reporting for rapes in 2013, the violent crime measure excludes incidents of rape (see https://ucr.fbi.gov/crime-in-the-u.s/2013/crime-in-the-u.s.-2013/violent-crime/rape); similarly, arsons are excluded from the property crime measure because it is inconsistently reported across counties (see https://ucr.fbi.gov/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/property-crime/arsonmain).

[40] Note that the census controls for New York City are calculated as the population-weighted average of the controls across the New York City boroughs, weighting by 2000 census population.

## County Level Demographic and Economic Variables

Below is a brief description of the sources used for construction of county level demographic and economic variables and, if applicable, the method of interpolation used to estimate data for any missing years. Table 2 below reports the data sources and years for which the data are available by variable category.

**Population Characteristics**

Data on annual population and population characteristics by county were downloaded from the Census Bureau's Population Estimates Program (PEP).[41] These data are available on an annual basis from 1990 to 2017.

**Education**

Data on the distribution of educational attainment by county from 1990 to 2017 were created from two sources: (1) decennial Census data from 1990 and 2000; and (2) the 5-year American Community Survey as compiled by the National Historical Geographic Information System (NHGIS).[42] To create an annual time series, data are interpolated for years not reported in either the Decennial Census or the 5-year American Community Survey data. Raw data from the decennial census is used as of the census year for 1990 and 2000. Where ACS 5-year estimates are available, the midpoint of each 5-year period is used as an estimate of the value of the variable for that mid-point year (e.g. 2005-2009 is used for 2007, 2006-2010 is used for 2008, etc.). Linear interpolation is used to fill in gaps between available data points. Following linear interpolation, the most recent data point is carried through to 2017. Based on these sources, the following education categories are constructed: less than high school; high school graduate (but not college); at least some college; college graduate or greater.

**Median Household Income**

Median household income data is obtained from the same sources and constructed in the same manner as the education statistics. Median household income is converted into 2010 dollars using annual BLS CPI data.[43] When used in regression analyses, median household income is expressed in thousands of dollars.

**Poverty Status**

Percentage of people living in poverty, by year, from 1980 to 2017 is constructed in a similar manner to the education and median household income time series. This variable is defined as the number of persons with income below poverty level divided by the total number of persons for whom poverty status is determined. The decennial census identifies people as below the poverty level based on their

---

[41] U.S. Census Bureau; Population Estimates Program: https://www.census.gov/programs-surveys/popest.html
[42] Steven Manson, Jonathan Schroeder, David Van Riper, and Steven Ruggles. *IPUMS National Historical Geographic Information System: Version 13.0* [Database]. Minneapolis: University of Minnesota. 2018. http://doi.org/10.18128/D050.V13.0
[43] Bureau of Labor Statistics, CPI-All Urban Consumers (Current Series): https://data.bls.gov/timeseries/CUUR0000SA0?output_view=pct_1mth

prior year income. The 5-year ACS surveys identify people whose past 12-month income was below poverty level (surveys are run continuously over the 5-year period).

**Urban Percentage of the Population**

This variable is defined as the number of people living in urban areas (including both urbanized areas and urban clusters) divided by the total number of people living in urban and rural areas. Urbanized areas and urban clusters are determined by the Census Bureau based on the population within a geographical area. This metric is available in the decennial Census data. Data in non-reporting years between 1990 and 2010 are estimated using linear interpolation between the available decennial census years. Data post 2010 is held constant and set at the 2010 level.

**Employment Statistics**

Data on the labor force were obtained from the BLS's Local Area Unemployment Statistics program.[44] Data are available annually at the county level from 1990 to 2017. The unemployment rate is calculated from the BLS data as the number of unemployed persons divided by the total labor force. The employment to population ratio is calculated using the BLS's total number of employed persons divided by the Census total adult population age 15+.

**Industry Characteristics**

Data on industry characteristics were obtained from the Census' County Business Pattern data.[45] Industry shares are calculated for each state or county as the number of employees in an industry divided by the number of employees in all classified industries (as identified by NAICS code). Data are available annually at the county level from 1990 to 2016.[46] 2017 values are held constant at the 2016 level.

[44] Bureau of Labor Statistics; Local Area Unemployment Statistics: https://www.bls.gov/lau/
[45] US Census Bureau; County Business Patterns: https://www.census.gov/programs-surveys/cbp.html
[46] Pre-1990 data are also available and are used for calculation of the 10-year change in manufacturing employment. Prior to 1986, SIC codes were used by County Business Patterns. An SIC to NAICS code lookup (https://www.census.gov/eos/www/naics/concordances/1987_SIC_to_2002_NAICS.xls) was used to identify relevant manufacturing industry codes.

CONFIDENTIAL

## Table 2

## Economic and Demographic Variables with Data Sources and Years Reported

| Year | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Population Characteristics**<br>- Population<br>- Percent Male/Female<br>- Percent <15, 15-29, 30-44, 45-64, 65+<br>- Percent White/Black/Other<br>- Percent Hispanic | 1990 PEP | 1991 PEP | 1992 PEP | 1993 PEP | 1994 PEP | 1995 PEP | 1996 PEP | 1997 PEP | 1998 PEP | 1999 PEP | 2000 PEP | 2001 PEP | 2002 PEP | 2003 PEP | 2004 PEP | 2005 PEP | 2006 PEP | 2007 PEP | 2008 PEP | 2009 PEP | 2010 PEP | 2011 PEP | 2012 PEP | 2013 PEP | 2014 PEP | 2015 PEP | 2016 PEP | 2017 PEP |
| **Education**<br>- Percent Less than High School<br>- Percent High School Only<br>- Percent Some College<br>- Percent College or Higher | 1990 Census | Interpolated | | | | | | | | | 2000 Census | Interpolated | | | | | | 2005-2009 ACS | 2006-2010 ACS | 2007-2011 ACS | 2008-2012 ACS | 2009-2013 ACS | 2010-2015 ACS | 2011-2016 ACS | 2012-2017 ACS | 2013-2017 ACS | | |
| **Employment Statistics**<br>- Employment to Population Ratio<br>- Unemployment Rate | 1990 BLS | 1991 BLS | 1992 BLS | 1993 BLS | 1994 BLS | 1995 BLS | 1996 BLS | 1997 BLS | 1998 BLS | 1999 BLS | 2000 BLS | 2001 BLS | 2002 BLS | 2003 BLS | 2004 BLS | 2005 BLS | 2006 BLS | 2007 BLS | 2008 BLS | 2009 BLS | 2010 BLS | 2011 BLS | 2012 BLS | 2013 BLS | 2014 BLS | 2015 BLS | 2016 BLS | 2017 BLS |
| **Median Household Income**<br>(Converted to 2010 Dollars using Annual CPI data) | 1990 Census | Interpolated | | | | | | | | | 2000 Census | Interpolated | | | | | | 2005-2009 ACS | 2006-2010 ACS | 2007-2011 ACS | 2008-2012 ACS | 2009-2013 ACS | 2010-2015 ACS | 2011-2016 ACS | 2012-2017 ACS | 2013-2017 ACS | | |
| **Poverty Rate** | 1990 Census | Interpolated | | | | | | | | | 2000 Census | Interpolated | | | | | | 2005-2009 ACS | 2006-2010 ACS | 2007-2011 ACS | 2008-2012 ACS | 2009-2013 ACS | 2010-2015 ACS | 2011-2016 ACS | 2012-2017 ACS | 2013-2017 ACS | | |
| **Percent Urban** | 1990 Census | Interpolated | | | | | | | | | 2000 Census | Interpolated | | | | | | | | | 2010 Census | | | | | | | |
| **Industry Chracteristics**<br>- Percent Agriculture/Mining/ Construction/Utilities<br>- Percent Manufacturing<br>- Percent Retail/Transportation<br>- Percent Professional Services<br>- Percent Health Care/Accomodation/ Food/Other Services | 1990 CBP | 1991 CBP | 1992 CBP | 1993 CBP | 1994 CBP | 1995 CBP | 1996 CBP | 1997 CBP | 1998 CBP | 1999 CBP | 2000 CBP | 2001 CBP | 2002 CBP | 2003 CBP | 2004 CBP | 2005 CBP | 2006 CBP | 2007 CBP | 2008 CBP | 2009 CBP | 2010 CBP | 2011 CBP | 2012 CBP | 2013 CBP | 2014 CBP | 2015 CBP | 2016 CBP | |

Notes: PEP: US Census Bureau Population Estimates Program
ACS: 5-year American Community Survey
Census: Decenial Census
BLS: Bureau of Labor Statistics Local Area Unemployment Statistics program.
CBP: County Business Patterns
When necessary, rates are calculated using relevent population from the Population Estimates Program.
Interpolated values are a linear interpolation between the preceding and following measured value.